DOCUMENT RESUME

ED 301 770 CE 051 595

AUTHOR Thomas, Ruth G.
TITLE The Tailored Response Test: An Approach to Assessing Higher Order Thinking Related to Work Roles and Contexts.
INSTITUTION Minnesota Univ., St. Paul. Minnesota Research and Development Center for Vocational Education.
SPONS AGENCY Minnesota State Board of Vocational-Technical Education, St. Paul.
PUB DATE Nov 88
NOTE 68p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Cognitive Measurement; *Decision Making Skills; Diagnostic Tests; Secondary Education; *Situational Tests; Student Evaluation; *Test Construction; Test Reliability; Test Validity; *Vocational Education; Work Environment
IDENTIFIERS *Tailored Response Test; *Thinking Skills

ABSTRACT

The Tailored Response Test (TRT) is an approach to assessment that presents information or real-life situations and assesses how the respondent interprets and handles information or situations. Respondents (N=127) were asked to respond to 12 test segments by editing a printed passage or other material following a video presentation of a situation or information in the field of child care. The edited paragraph or material was then hand scored using a scoring template or scored more quickly by using a microcomputer. A study examined the potential of the TRT as an approach to assessing higher-order thinking and knowledge use within a specific domain and practice area relevant to vocational education. Results of the pilot tests conducted to assess the TRT indicated that it successfully focuses on the totality of situations rather than on only one or two aspects of a situation, reflects expertise in an identified body of knowledge, avoids the potential for test-wiseness being reflected in responses, and is reliable and valid at levels accepted for classroom use, program evaluation, and potential development as a standardized test. It was thus deemed to be a promising approach to determining contributions of vocational education to developing higher-order thinking skills in students that is relevant to work roles and contexts. (Information on the content/practice domain for the TRT is appended.) (MN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED301770

Minnesota Research and Development Center for Vocational Education

The Tailored Response Test: An Approach to Assessing Higher Order Thinking Related to Work Roles and Contexts

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D Pucel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Minnesota Research and Development Center
for Vocational Education
Department of Vocational and Technical Education
College of Education
University of Minnesota
St. Paul, Minnesota 55108

CE 051 575

The Tailored Response Test: An Approach to Assessing Higher Order Thinking Related to Work Roles and Contexts

BY

Ruth G. Thomas

November 1988

Minnesota Research and Development Center for Vocational Education
Department of Vocational and Technical Education
R460 VoTech Building, 1954 Buford Avenue
University of Minnesota
St. Paul, Minnesota 55108

© Copyright 1988, MRDC

Funding for this project was provided by the State Board of Vocational Technical Education, the Minnesota Department of Education, and the Department of Vocational and Technical Education, University of Minnesota.

Interpretations of the findings described in this report represent those of the authors, and are not necessarily those of the State Board of Vocational Technical Education, the Minnesota Department of Education, or the University of Minnesota.

The University of Minnesota is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, religion, color, sex, national origin, handicap, age, veteran status, or sexual orientation.

ACKNOWLEDGMENTS

The research reported in this monograph could not have been completed without the generous contributions of time, effort, resources and expertise from a number of people and institutions. Sincere gratitude is expressed to the following for their willing participation in the research:

- The six individuals who served on the panel of experts and their institutions for time that numerous reviews of project materials took and for invaluable input in the development of the experimental tailored response test:

  Dr. Betty Cooke, Research Associate in the Home Economics Education Division at the University of Minnesota and formerly a parent educator in several Twin Cities parent education programs

  Mary Ann Curtain, Director, Child Development Program, Anoka Technical Institute

  Dr. Martha Farrell Erickson, Coordinator of Project STEEP, Department of Educational Psychology, University of Minnesota

  Lynn Galle, Coordinator and Lecturer, Institute of Child Development, University of Minnesota

  Carol Morgaine, doctoral student in the Home Economics Education Division at the University of Minnesota and formerly director of the early education program at Bethel College

  Richard Tvedten, Director, Child Development Program, Northeast Metro Technical Institute 916

- The instructors who allowed invasion of their nursery school setting with cameras, lights and extension cords during the production of approximately eight hours of video tape:

  Ann Mooney-McLoone, Instructor, Institute of Child Development, University of Minnesota

  Barb Murphy, Early Childhood Specialist, Institute of Child Development, University of Minnesota

- University Media Resources for the grant which provided the personnel and equipment that produced professional level video-tape test

stimulus material and the following individuals who were instrumental in the funding and production of the tape:

Arnold Walker, Associate Director for Television, for his assistance in obtaining the grant and his encouragement of and enthusiasm for the project.

Earl Niemi, Director and Producer, for his direction and management of the video-tape production.

- The Center for Early Education (CEED) for access to video tapes owned by the Center that contained appropriate stimulus material for the tailored response test, and especially Dr. Richard Weinberg, CEED Director, and Erna Fishhaut, CEED Coordinator, for their willingness to share the video tapes and their interest in and enthusiasm and support for the project.

- Carol Chamberlin, Coordinator, Child Development Program, Santa Monica College, Santa Monica, California, for her permission to include portions of a commercial video tape she has produced in the experimental tailored response test.

- Staff at the University Child Care Center, especially Patty Finstad, for willingness to allow use of the Center site for preliminary observation of situations as input to the planning of video-tape production.

- Child development, child care, and early education instructors for sharing their class time and their students so that essential student data could be obtained for the production of the tailored response test:

Ann Mooney-McLoone, Instructor, Institute of Child Development, University of Minnesota

Ellen Werr, Child Development Instructor, Suburban Hennepin Technical Institute, North Campus

- Instructors, programs and institutions who provided the opportunity to pilot test the tailored response test with their students:

Sue Snicale, Child Development Instructor, Blaine High School

Mary Ann Curtain, Director, Child Development Program and Debbie Compton, Instructor, Evening

Child Development Program, Anoka Technical Institute

Richard Tvedten, Director, Child Development Program, Northeast Metro Technical Institute 916

Pat Koester, Director, Child Development Program, Suburban Hennepin Technical Institute, South Campus

Darlene Hallberg, Instructor, Human Relations Program, Inver Hills Community College

Vivian Houk, Instructor, Early Education Program, Bethel College

Sue Foster, Director, Early Education Program, Bethel College

Beth Cutting, Instructor, Home Economics Education, University of Minnesota

Bob Weigand, Instructor, Institute of Child Development, University of Minnesota

Barb Murphy, Instructor, Institute of Child Development, University of Minnesota

Thanks are also due to Gale Smith, Graduate Assistant for the project, who contributed to a wide range of technical aspects of the project, to the data gathering and assembling processes, and who developed the initial draft of the content/practice domain. Especially noteworthy contributions were extensive hours spent editing and re-editing miles of video tape, extensive effort in pilot test data collection, production of scoring templates, assembling of test packets, coding and scoring the tests, and computerizing and verifying data for analysis. Faye Caskey's efforts in the data verification process are also gratefully acknowledged.

Several individuals provided assistance related to data analysis. Bruce Center's and Mark Davison's efforts in trouble shooting a non-functioning statistical program were critical at a crucial stage in the project and Jack Merwin provided useful recommendations.

The Minnesota Research and Development staff, especially Nora Gille, Bonnie Henderson and Kimberly Pippel, scheduled and arranged for panel of expert meetings throughout the year, sent out correspondence to the myriad of individuals involved in this project who are listed above, kept the project budget straight, did final typing revisions on the manuscript, and oversaw those final details that always accompany the

publication of a report. Their assistance in all of these tasks is gratefully acknowledged.

Thanks is also due to Dr. David Pucel, Director of the Minnesota Research and Development Center, for his positive responses to many requests which made the project go better, for the flexibility he demonstrated in facilitating opportunities and problem solving during the course of the project and for his support of the project within the State and University contexts.

Finally, thanks to the State Board for Vocational Education for providing the funds to make this project possible and for their recognition of the need for assessment tools that enable documentation of accomplishment of significant educational goals. Special thanks to Bill Stock, project monitor and State Board for Vocational Education staff member, for his interest in the project and his commitment to educational improvement.

TABLE OF CONTENTS

	Page
Abstract	viii
Chapter 1: Introduction, Problem, Research Purposes, Background and Educational Significance of the Study	1
Introduction: Higher Order Thinking Research Program.	1
Problem and Research Purposes	2
Theoretical and Conceptual Background Underlying the Tailored Response Test.	6
Educational Significance of the Study	9
Chapter 2: Research Methods and Procedures.	11
TRT Production Procedures	11
TRT Pilot Testing Procedures.	19
Test Scoring Procedures	20
Data Analysis	20
Chapter 3: Pilot Test Results, Summary, Conclusions, Implications and Recommendations	23
Reliability	23
Internal Validity	24
External Validity	26
Usability	29
Summary, Conclusions and Implications	31
Recommendations	34
References	37
Appendix	39

List of Tables

Table 1:	Readability of TRT Paragraphs	18
Table 2:	Internal Consistency Reliability of TRT Paragraphs (N=127).	23
Table 3:	Pearson r Correlations Indicating Degree of Relationship Between Test Segments (N=113-127)	25
Table 4:	Relationship Between Segment Scores and Total Score (N=113)	25

List of Tables (continued)

Table 5: Relationship of Total Score and External Variables 27
Table 6: Comparison of Total Scores by Educational Level (N=113) 28

ABSTRACT

A problem that all subject areas within vocational education face is the lack of adequate assessment tools for documenting learning of higher cognitive processes and knowledge that support work-role-related expertise. Assessment of learning is typically limited to paper and pencil objective tests that mostly measure knowledge recall and observational check lists of work performance that have questionable reliability and validity and that focus on observable behavior rather than on intellectual structures and processes that guide behavior. Consequently, the ability of vocational educators to help learners diagnose their strengths and weaknesses and improve their capabilities has been largely limited to knowledge storage issues and physically observable aspects of performance. Existing approaches that have been used to assess thinking have practical and other types of limitations. Tests designed to assess thinking are concerned with general cognitive abilities - not thinking within a specific knowledge domain. Simulation tests are costly due to expensive development requirements and expenses involved in the use of experts to judge the performance of each person tested. Essay tests are impractical to use often or with large groups of students because scoring is time consuming. Further, because scoring of essay tests is also imprecise, the reliability of these tests is questioned.

The purpose of the research reported here was to investigate the tailored response test (TRT) as a potential approach to assessing learning of perceptual and thinking processes, including complex judgment processes, relevant to work roles. The TRT is an approach to assessment that more closely simulates the actual functions and operations to be predicted than is possible using traditional written, objectively scored multiple choice tests. The TRT is intended to provide test responses that are more purely a function of the knowledge of a field or subject and less a function of alertness to surface cues provided by traditional types of test items as to a "right" answer. At the same time, the TRT is easily, quickly, and precisely scored. The TRT is applicable to a knowledge discipline content area or to an area of practice.

The TRT presents information or real-life situations and assesses how the respondent interprets and handles information or situations. This is accomplished by asking the respondent to edit a printed passage or other material following presentation of the situation or information. The edited paragraph or material is then hand scored using a scoring template or can be even more quickly scored using a microcomputer.

A TRT was developed and pilot tested in the area of supporting children's social development. Reliability and validity data suggest that the TRT is reliable and valid at levels acceptable for classroom use and for potential development as a standardized test. Results indicate that the TRT is a promising approach to the assessment of learning of perceptual and thinking processes, including complex judgment processes relevant to work roles. Further, the TRT proved to be usable and practical for classroom teaching situations. The TRT appears to be transferable to a wide range of work roles and programs in vocational education and is adaptable to interactive computer formats, including video disk, and to satellite transmission. The TRT development process has been closely documented in this study and is now at a point where a number of researchers can conduct studies using the approach across areas of vocational education. The TRT approach to testing is anticipated to be useful in training programs for establishment of training requirements, identification of training problems, diagnosing student learning needs, development of training programs, and evaluation for the effectiveness of training programs.

CHAPTER 1

INTRODUCTION, PROBLEM, RESEARCH PURPOSES, BACKGROUND AND EDUCATIONAL SIGNIFICANCE OF THE STUDY

Introduction: Higher Order Thinking Research Program

The research reported in this monograph was conducted as part of the Higher Order Thinking Research Program at the Minnesota Research and Development Center for Vocational and Technical Education located at the University of Minnesota. The purposes of the Higher Order Thinking program of research are to conduct research on:

- the nature of problems requiring higher order thinking that are of concern in vocational education
- the nature of mental processes and structures that underlie expertise in specific knowledge domains related to work roles and contexts
- instructional design for developing, facilitating and improving mental processes and structures associated with specific knowledge domain expertise
- assessment of mental processes and structures underlying expertise in specific knowledge domains related to work roles and contexts

It is intended that this research will result in a better understanding of the nature of problems and expertise in areas relevant to vocational education and better ways of developing and assessing that expertise. This monograph reports results of research focused on the fourth purpose, assessment of thinking processes and mental structures underlying expertise in specific knowledge domains related to work roles and contexts.

The focus of the research in the Higher Order Thinking Research Program is on understanding, developing and assessing mental processes and structures underlying expertise in specific knowledge domains. A specific knowledge domain is what one needs to know in order to successfully practice in a particular profession, trade or role. A specific knowledge domain is acquired by education, training and experience which exposes the individual to the concepts, principles, processes, and phenomena in a domain of knowledge (e.g., diseases, electronics, children, crops, accounts, etc.).

Mental processes include the processing of information as it is influenced by perception and individual dispositions. Mental structures include the forms, organization, arrangements and systems in which knowledge exists in the human mind. Expertise refers to the possession of a high level of skill or proficiency in identifying and solving a problem, resolving a situation or performing some function. As it is used here, it also refers to the production of particularly creative, interesting or insightful thought, the noting of especially subtle nuances. Expertise is demonstrated by depth and breadth of comprehension and interpretation, and by the quality of conclusions, evaluations and judgments. A more complete discussion of terms and concepts central to this research program can be reviewed in Thomas and Litowitz (1986).

The Higher Order Thinking Research Program has been funded since 1985 by the Minnesota State Board for Vocational Education. In the first year of the program, an agenda for inquiry investigating higher order thinking in relation to vocational education was produced (Thomas, R. & Litowitz, L., 1986). Two studies responding to the portion of the inquiry agenda calling for research that contributes to understanding of the nature of problem solvers were completed during the Higher Order Thinking Research Program's second year. These two studies focused on answering the question, "What knowledge and mental processes guide, organize and form effective actions in working with specific knowledge domain problems?" The studies, reported in a forthcoming monograph (Thomas, 1988), examined relationships between knowledge and mental processes and the ability to solve problems.

The tailored response test (TRT) project reported in the present monograph was completed in the third year of the research program. This project addressed the need for tools to assess thinking processes and mental structures underlying expertise in specific knowledge domains related to work contexts.

In the forthcoming fourth year of the research program, instructional designs for developing thinking processes and knowledge structures underlying expertise will be developed and tested.

Problem and Research Purposes

A problem that all subject areas within vocational education face is the lack of adequate assessment tools for documenting learning of work-role-related higher cognitive processes and the knowledge that supports them. Vocational educators claim they help students learn complex skills that require a sophisticated level of knowledge, skill, and cognitive functioning, but assessment of learning is typically limited to paper and pencil objective tests that mostly measure knowledge

memorization and recall. This problem is not unique to vocational education and pervades all subject areas and educational levels.

The need for assessment tools that can reliably and validly assess learners' ability to make interpretations and judgments in situations like those they will encounter in real life is well recognized. Nevertheless, more progress has been made in applying new research findings regarding higher order thinking to instructional design than to assessment. Few tested approaches exist for assessing decision making and problem solving processes associated with complex, nonroutine, ill-defined problem situations within specific knowledge and practice domains. This poses a problem in evaluating new instructional designs purported to teach thinking processes.

Assessment approaches that attempt to go beyond knowledge recall, to knowledge application, are currently limited to such devices as observational checklists, interpretive exercise multiple choice items, essay tests and simulations. Checklists often have unknown or questionable reliability and they focus on performance behaviors rather than the mental processes and structures underlying performance. While checklists document the presence or absence of performance behaviors, they do not point to the reasons behaviors may be demonstrated or missing. At best, checklists are gross indicators which leave a need for more precise and refined assessment tools that can tap the subtle, covert differences in perceptions, knowledge and thinking processes that underlie expertise.

The importance of focusing on the mental processes and structures underlying expertise is that these are the key elements that drive action and differentiate high quality performance from lower quality performance. Research within vocational education substantiates this claim (Thomas, 1988) as does research throughout the cognitive science literature. While some existing alternatives to assessing thinking mentioned above can focus on the mental processes and structures, they exhibit other problems.

For example, the interpretive exercise multiple choice test does present the respondent with information, a problem, puzzle, dilemma, or other situation about which thinking is required, but the objective responses provided in this type of test often offer clues to the test-wise responder. A test-wise individual's response may be more an indication of their ability to detect such clues than a true indicator of their mental structures and thinking processes. Testing of thinking processes requires that the test materials not present overly defined situations or problems because an important aspect of what is to be tested is how individuals view and interpret a situation, what they focus on and don't notice, and how they see

or represent a problem. The response choices in the interpretive exercise "give away" what the problem is to a certain extent and reduce the validity of the response as a true measure of the respondents' capability in the knowledge area.

Essay tests provide another means for assessing higher order thinking but are impractical to use often or with large groups of students because scoring is time consuming and is subject to low reliability both over time within a single rater and across raters.

Simulation tests such as the in-basket test (Frederiksen, Saunders & Wand, 1957) or tests using equipment simulators have typically been limited in their practicality due to high initial development costs and expenses involved in scoring procedures that use experts to judge the performance of each person tested. At the same time, simulation tests are particularly well suited for presenting situations to the learner with all the complexity associated with real-life situations.

The need for more sophisticated assessment tools that are also practical for use in educational situations does not only exist at the end of the learning process. Pre-instructional diagnosis of learner strengths and weaknesses has also been largely limited to knowledge storage and physically observable aspects of performance. This situation has seriously limited educators' abilities to more directly and precisely affect the development of mental processes and structures that guide and form expertise in the learning process.

Finally, the current nationwide dilemma concerning teacher testing reflects a growing general concern with reliable and valid documentation of students' capabilities as they enter training programs, with more rigorously documented certification of their capabilities as they exit training programs, and with improved prediction of their capabilities in relation to personnel selection decisions.

The increasing needs and pressures for assessments of complex thinking processes are fueling an interest in solving the reliability, validity, practicality and cost problems associated with current approaches to testing higher level thinking. The research reported in this monograph contributes to addressing that need. The purpose of the project was to investigate the potential of the tailored response test (TRT), an experimental approach under investigation at Educational Testing Service's (ETS) Center for Occupational and Professional Assessment, as an approach to assessing higher order thinking and knowledge use within a specific knowledge domain and area of practice. Specifically, the TRT was investigated for its potential to:

- assess student learning of perceptual and interpretive thinking processes and complex judgment processes relevant to expertise in a work role requiring knowledge in a specific domain

- go beyond assessing possession of knowledge to assessing the ability to use knowledge

Objectives of the study were to develop and evaluate the TRT as an approach to assessing learning within a specific knowledge domain and practice area relevant to vocational education. Criteria for evaluating the TRT are listed below along with the rationale for why they were chosen:

1. Focus on the totality of situations rather than on only one or two aspects. Rationale: One of the problems in the transfer between situations used as examples or trials in educational processes and real life is that educational situations have been simplified, presenting the learner with only selected aspects of a situation or problem. Real-life problems confront the learner with all aspects all at once in a barrage of complexity and confusion that must be sorted out by the learner. Tests that present the learner with an unrealistically narrow range of concurrent factors are likely to have little fidelity with real situations, and little power in predicting the learner's capabilities in a real situation.

2. Focus on higher level thinking processes. Rationale: Few testing approaches address thinking processes. Currently available testing technologies emphasize knowledge storage or overt performance, ignoring the underlying mental processing of perceptions and information. Yet, research has shown that the underlying mental processing differentiates experts from novices. Testing typically focuses more on what can be tested within current technologies rather than on what needs to be tested. Testing too often involves tools that are practical to use and relatively inexpensive to develop, but which measure mostly knowledge recall. If education is to directly and effectively affect mental processing, such processing will need to be assessed for diagnostic and evaluative purposes.

3. Responses reflect expertise in an identified body of knowledge. Rationale: To date, existing tests of thinking processes focus on general thinking ability. Existing tests do not detect the integration of knowledge with thinking processes that experts in a particular field display.

4. Can be administered within the time structure of typical classrooms. Rationale: While there is a need for standardized tests to be developed that can certify and

predict higher order thinking within specific knowledge and practice domains, an immediate need is for tests that can help instructional developers and teachers determine the effectiveness of new instructional designs aimed at teaching thinking in classrooms.

5. Easily and reliably scorable in a short amount of time. Rationale: The essay test can be used to assess thinking processes if scorability is not an issue and if reliability is not a paramount concern. Few educational situations provide the luxury of time investment that scoring of essay tests requires. The care needed to achieve scoring reliability on such tests is difficult to consistently apply. An assessment approach is needed which will allow more frequent and larger scale testing of thinking processes than is practical using essay tests. No test is likely to be extensively used by classroom teachers unless it meets this criterion.

6. Assumes that real life problems don't have one right answer but rather, have several possible resolutions which may be appropriate and none of which may be perfect. Rationale: A serious limitation of objective type tests is that only one answer is "right." Real-life solutions to complex problems tend to be "better" or "worse" rather than right or wrong.

7. Avoid potential for test score to reflect "test-wiseness" - the ability to guess the right answer from surface characteristics of the items. Rationale: It is well known that people differ in their skill at test taking. Objective tests are particularly subject to displaying cues regarding the right answer to the skillful test taker. Consequently, test scores of individuals who are skilled at taking objective tests do not solely reflect their expertise in a content or practice domain.

8. Is reliable and valid at levels acceptable for classroom use, program evaluation, and for potential standardization. Rationale: Reliability and validity are basic to any measurement tool and are perhaps the most challenging criteria to meet in a tool intended to measure complex thinking.

Theoretical and Conceptual Background Underlying the Tailored Response Test

In order to test thinking at higher levels rather than knowledge recall, a test must confront the individual with a problem, puzzle, dilemma, or other situation that requires the learner to do one or more of the following: acquire and interpret information; form conclusions, judgments, strategies. A more complete discussion of these processes and evidence regarding their nature is provided in earlier publications (Thomas & Litowitz, 1986; Thomas, 1988).

Testing of thinking within specific knowledge domains, particularly the testing of thinking in practice-oriented situations, must present the respondent with situations from the practice (work role) context that require both the thinking processes described above and domain knowledge. Further, the situations must capture the complexity of real-life situations produced by the need to consider many different factors simultaneously.

Because simulations appear to offer particular promise for valid testing of thinking in knowledge-domain-related practice situations due to their ability to achieve a high degree of fidelity with real-life situations, ways of reducing their current practical limitations are worthy of investigation. The TRT was identified by Thornton (1988) as an approach to assessment that more closely simulates the actual functions and operations to be predicted than is possible using traditional written, objectively scored multiple choice tests. Because the TRT was also identified as reducing the practicality and cost limitations of the in-basket and other simulation type tests (Thornton, 1988), it appeared to be worthy of in-depth investigation as a potential tool for assessing complex thinking and knowledge structures relevant to specific knowledge and practice domain expertise of interest in vocational education.

The TRT presents simulations or records of real-life situations via video tape, audio tape, or printed material and assesses how the respondent views the situation. This is accomplished by asking the respondent to edit a printed passage or other material that discusses or refers to the situation. Respondents are asked to edit or correct a paragraph, form, graph, or table in light of their judgments about the situation by crossing out words, phrases, symbols or sentences. The final edited version of the response material represents the respondent's judgment about the correctness of actions taken in a given situation, possible future actions, and attitudes displayed by individuals in the simulation. In addition to evaluative judgments, responses on the TRT reflect perception of cues and interpretation of conditions and events. Because there are no response choices indicated or any indication of where editing is to occur in the response material, it is possible to avoid cuing the respondent as to what the problem is and where a response is called for. This feature enables the TRT to generate responses that are more purely a function of the knowledge of a field or subject and less a function of alertness to test item surface cues to the "right" answer.

Because the TRT allows scoring of responses using a microcomputer, costly scoring procedures typically used in simulation type tests for open-ended responses (such as essays) are avoided. Further, in classroom situations, the teacher has

control over the scoring process for the TRT rather than being required to send the test to expert scorers.

The TRT is applicable to a knowledge discipline content area or a professional practice achievement area (Thornton, 1988). While the TRT can be applied to both occupational or scholastic content or achievement areas, early work on the test at ETS has focused on occupational areas. Initial work with the TRT at ETS has involved assessing operating room nurses and police officers. More recent TRTs have been developed at ETS to assess electronic trouble shooters and navy electronic equipment operators.

The roots of the TRT come from the Cloze test (Taylor, 1953) which was originated to assess reading comprehension. An intermediate step in the evolution of the TRT was the cloze-edit procedure suggested by Manning (1985). Extensive adaptation of the cloze-edit procedure by Richard Thornton (1988) resulted in the first experimental TRT produced at ETS.

Early work with the TRT at ETS suggests that this assessment approach can discern students' thinking in nonroutine work situations where appropriate action is not automatic or totally rule-based but must be determined through consideration of many factors, alternatives and potential consequences. Further, the early work is promising regarding the TRT's reliability and validity. Thornton (1988) reported reliability coefficients on TRT scores over 131 items of .814 (Kuder Richardson #20) with a standard error of 3.56. Validity data reported by Thornton (1988) consisted of correlation coefficients of association between TRT scores of 100 subjects and four external variables. These coefficients were:

TRT with an open-ended response simulation: .523
TRT with candidate selection test score: .345
TRT with grade point average at end of training program: .412
TRT with a written work knowledge test: .310

All coefficients were significant at the .01 level.

Because the range of scores on the external criterion variables was considerably narrowed due to the subjects having been selected into a program based on their attainment of a minimum score level on a selection test, a procedure was used to correct for range restriction (Thorndike, 1949, pp. 173-174). Correlation coefficients of association between the TRT scores and the corrected scores for the four external variables are reported by Thornton as follows:

TRT with an open-ended response simulation: .749
TRT with candidate selection test score: .749

TRT with grade point average at end of training program: .759
TRT with a written work knowledge test: .536

Based on evidence available regarding the TRT, it was concluded that this approach was highly promising as a way of assessing higher order thinking within specific knowledge domains of interest and concern in vocational education. The project reported here was undertaken to test that conclusion in relation to one such knowledge domain.

Educational Significance of the Study

A testing approach which assesses higher order thinking and is usable in instructional settings would have considerable value in today's environment which is increasingly demanding evidence that schools are developing students' thinking processes. Such a testing approach would also be useful in evaluating alternative instructional designs for effectiveness in developing thinking processes.

It is anticipated that the TRT approach to testing may be useful in training programs for establishment of training requirements, identification of training problems, diagnosing student learning needs, development of training programs, and evaluation of the effectiveness of training programs. Further, it is anticipated that the TRT may predict effectiveness in work situations related to the TRT test items. Consequently, it may eventually find use in selection, classification, and certification of personnel. The TRT is usable as a classroom test or it can be developed as a standardized test. The possibility of administering the TRT by satellite transmission exists and would be advantageous for large scale testing programs or testing programs which must occur over wide geographic areas and yet be centrally controlled. The TRT is adaptable to interactive computer formats, including videodisc. Such adaptation has already been done at ETS.

CHAPTER 2

RESEARCH METHODS AND PROCEDURES

Methodology used at ETS to develop experimental TRTs was adapted for the production of the TRT investigated in this study. The following sections report the procedures that were used to produce the experimental TRT assessment device and to pilot test it.

TRT Production Procedures

1. Establish the content/practice domain: The content/practice domain chosen for the focus of the experimental TRT was supporting young children's social development. This content/practice domain was chosen for several reasons. First, it is an area having considerable complexity and imprecision in the judgments it requires of adults who work with and care for children. It was assumed that a knowledge domain characterized by complexity and imprecision in its application would provide a stringent test of the ability of the TRT to produce reliable scores. Second, there is a substantial, rigorous knowledge base regarding this developmental area. Consequently, it was anticipated that this extensive and substantive content/practice domain would be a rich resource for building test items. Third, difficulties have prevailed in training programs in assessing individuals with respect to this domain using traditional assessment methods. Consequently, this domain was perceived to exemplify characteristics of many similar domains for which the TRT might be particularly applicable. Fourth, since it would be necessary to identify or develop stimulus materials focused on the content/practice domain, the choice had to be an area where such material was already available or could be readily developed. Because social development is an aspect of most interchanges between child care and early education practitioners and children, and because social development pervades so much of what goes on in the daily routines of day care and early education settings, it was anticipated that stimulus material in this area could be readily obtained or developed. Finally, this knowledge area was one in which the investigator had background. This was thought to be especially important in the item writing process.

   The knowledge domain is referred to in the above discussion as a content/practice domain to reflect the

linkage between knowledge of children and knowledge of adult actions relevant to goals for children's development. The concept of practitioners' knowledge being formulated and stored in terms of practice and action is central to the content/practice domain that was developed. While many test development procedures refer to a content domain, the term, content/practice domain, was intentionally used here.

Several types of materials formed the basis for development of the content/practice domain. Several published child care and early education curricula were reviewed. In addition, the aims, goals, and objectives of several local child development training programs were reviewed to identify the areas of competence, skill and personal qualities upon which they focused. Major areas of consistency were identified from this review. Research literature regarding adult roles, goals and practices that are supportive of children's social development and several child development texts in the area of social development were also reviewed to further identify areas and potential ways of structuring the domain. A review of a new text (Kostelnik, Stein, Whiren & Soderman, 1988) specifically devoted to the concept of supporting children's social development and containing consistent but more extensive and detailed information than the other sources was used as the primary basis for detailed development of the content/ practice domain. An important feature of this text was the provision of implications for practitioners and parents along with knowledge about children's social development. This text was written from the perspective of the practice of guiding children's social development. That perspective plus its up-to-date content made this text an especially useful source.

The structure of the content/practice domain that was developed had three levels (See Appendix). The most general level was called a function. A function was a special duty or performance requirement of a person in the course of their work. A function reflected an overall, broad goal related to children's social development. The second level was called an element. Elements were defined as subgoals within an area of social development or within a special duty or performance requirement. Elements are more action oriented than functions but are not observable without further specification and definition. Elements were identified within functions. The third level was called tasks. Tasks were what adults needed to do (acts) to accomplish the subgoals identified with elements. Tasks were identified within elements and were specified at an observable level.

The intent in establishing this structure for the content/practice domain was to avoid becoming mired down in observable tasks and risk losing sight of the meaning, significance and purpose of the tasks. The cognitive science literature indicates that mental structures which facilitate problem solving link acts to goals (Larkin, 1979). An attempt was made in the content/practice domain to clarify the social development subgoals that observable acts support. In this way, the content/practice domain was constructed to emphasize the relationship between knowledge and practice rather than depicting either only "knowing" or only behavior or action. This "knowledge embedded in practice" orientation of the content domain was especially important to the ability of the TRT that was produced to test respondents' interpretations and judgments regarding connections between goals and actions.

Once the content/practice domain was developed, it was reviewed and validated by the panel of experts described in the next section. This group's suggestions for changes were incorporated in the final version of the content/practice domain which is presented in the Appendix.

2. Assemble panel of experts: During the development of the content/practice domain, a second process was initiated. Individuals who were experts in working with children as indicated by their current professional position and recommendations from others were identified. Seven of these content and practice experts from the fields of child care, early education and child development were invited to serve on the panel of experts. These experts were on the teaching staffs of technical institutes, colleges and universities and had extensive experience in working with children and directing others in working with children. These experts were informed that their participation in the project would require four to five days of their time.

3. Identify critical incidents that reflect functions and elements of the content/practice domain: After reviewing the content/practice domain, the panel of experts was asked to identify incidents and situations that would be especially likely to reflect the functions and elements in the domain and that would typically be experienced by caregivers of children in day care and early education settings. Activities identified included transitions, small group activities, large group activities, routines such as lunch and nap, art, free play, one-to-one incidents where caregiver and a child interact, and times in which special stressors are present such as inadequate or inappropriate physical and staffing environments. Other factors identified as being important to consider in the selection of critical incidents

were the frequency of their occurrence, their significance to children's development and the seriousness of their consequences. Still further factors included the relevance of the incident for both parents and child care personnel, the complexity and challenge of the incident represented by its having a number of potentially appropriate solutions, and incidents which would involve preschool children of different ages.

Although the critical incident technique was not used per se, several features of this technique were reviewed and considered in relation to selection of critical incidents (Compton & Hall, 1972; Flanagan, 1954; Gilliom, 1977; Good, 1972; Miles, M. & Huberman, A., 1984; Witkin, B., 1984).

4. Develop, identify stimuli for critical incidents: Videotaped situations depicting the types of critical incidents discussed in the previous section were identified as the most appropriate format and medium for presentation of the TRT test stimulus material. A simulation test had been identified as appropriate for meeting requirements and solving assessment problems described in Chapter 1. Videotaped situations with adults and children were seen as vicarious simulation situations that would elicit projection onto and identification with the adult(s) in the situation and activate the mental processes and structures that actually being in similar situations would activate.

Further, the cues that enable interpretation of the critical aspects of adult-child situations are typically both visual (e.g., facial expression) and auditory (e.g., voice inflection). Video tape presents both visual and auditory cues simultaneously and with the complexity that real-life situations with children present. Adults who work with children must take from a barrage of stimuli those specific cues that will trigger interpretations which will enable the adult to act in a manner that is supportive of children's social development. Stimuli that do not present this barrage of stimuli which force educated selection of cues do not present the degree of challenge offered by real-life situations.

Development of stimulus materials for a simulation type test involving critical incidents often involves writing scripts, hiring actors and filming "scenes." Such a process allows the test developer a large degree of control over the incident content and flow. Such procedures also contribute to the high cost of test development. Situations with young children are difficult to stage in precisely this manner. Since activity of young children is typically intense and changes quickly, there is a high probability that situations of interest will occur frequently in normal day-to-day

occurrences. It was decided to video tape normal activities on two different days in two different situations involving different children and adults. Arrangements were made to do eight hours of video taping in a local nursery school setting. The only staging that was done was placing in the children's area certain articles likely to arouse curiosity and interest in children and lead to goals, discussion and actions related to specific areas of the content/practice domain.

In addition to the video taping described above, it was discovered that similar video tapes had recently been created at local group and family day care settings by another University unit that was willing to provide access to this material. Further, another video tape designed to teach aspects of children's social development and containing relevant critical incidents was discovered and permission granted by the producer to use portions of the tape in the research. These three sources of video tape yielded approximately seventeen hours of video tape in seven different settings and including approximately 30 different adults engaged in working with children.

Once obtained, each tape was segmented by episodes. An episode was a situation that had a focus on an activity, an incident or a goal. Some episodes had a beginning, a middle and an end. Some episodes contained a stream of activity rather than a situation with a clear beginning and end. Segments ranged from a few seconds to as many as 12 or 13 minutes in length. The total number of segments that resulted from the segmenting process was 107.

Each segment was then analyzed to determine which parts of the content/practice domain were reflected in the segment. The analysis was completed by two project staff members who were also members of the panel of experts. Rates of agreement were calculated. Agreement rates for the segments chosen for pilot testing ranged from a low of 75.25% to a high of 91.75%. The mean percent agreement over these segments was 84.77%.

Thirty-three of the 107 segments were chosen for possible inclusion in the test on the basis of content/ practice domain coverage, clarity and focus of the situation, and technical quality (visual and auditory clarity and focus).

5. Obtain panel of expert responses to critical incident stimuli: The panel of experts met as a group to view the 33 segments and arrive at consensus on the following: (a) the segment reflects significant aspects of the content/practice domain, (b) appropriate actions taken by the adult in the

situation, (c) inappropriate actions taken by the adult, (d) further actions that could be taken or that should have been taken by the adult, and (e) common errors made by students and practitioners in such a situation. Based on this input from the panel of experts, 23 segments were chosen from the 33 for further consideration.

6. Obtain open-ended responses to stimuli from sample of test respondent population: The 23 segments were shown to 42 technical institute and collegiate child care and early education students. Students were told that they were to focus on children's social development while viewing the segments and were asked to identify the following in writing for each of the segments: (a) appropriate actions taken by the adult in the situation, (b) inappropriate actions taken by the adult, and (c) further actions that could be taken or that should have been taken by the adult.

7. Panel of experts weights students responses to stimuli: The student responses obtained in the previous step were assembled and sent to the panel of experts individually. Each expert was asked to rate all student responses on two dimensions: First, yes or no, did the actions identified by the student sample occur in the segment or not and second, was the student response appropriate (+), irrelevant (0) or inappropriate (-). Frequency distributions of the panel of expert ratings were developed to assess the degree of agreement and disagreement among panel members' ratings.

8. Develop draft tailored response test paragraphs and scoring key based on panel of expert and open-ended responses: The rated student responses obtained in steps 6 and 7 and the panel of expert responses to the video-tape segments obtained in step 5 formed the basis for the construction of paragraphs. Paragraph length varied from one-third to three-fourths page, double spaced, depending on the amount and complexity of the content of the video segment. The paragraphs included statements about actions that were or were not taken by the adult in the video-taped situation, the appropriateness of the actions taken by the adult and actions that could potentially be taken. Some of the statements and phrases were reflections of panel of expert consensus on what was occurring in the situation and what was appropriate, inappropriate, and further action and some were statements that had been generated by the student sample that were deemed inappropriate by the panel of experts. The result was a paragraph which included material that should remain in the paragraph and material that should be deleted in order for the paragraph to represent expert judgment about the situation.

The initial draft of the paragraph was then analyzed twice. The first analysis was done to explicitly identify, modify, and add items. An item is a content-anchored word, phrase or sentence (R. Thornton, personal communication, January, 1988). The second analysis was done to structure each sentence in such a way that each word or phrase that constituted an item could be edited out or left in without grammatically affecting other items.

A duplicate copy of the paragraph was then developed into a scoring key. In the space above each line of the double spaced, duplicated test paragraph, items were identified and numbered. The items to be crossed out were then underlined on the scoring key.

9. Administer draft test to panel of experts; have panel of experts discuss test materials and answer key to identify problems and needed changes: The draft tailored response test was administered to the panel of experts who edited the paragraphs using the same procedure that test respondents would be expected to use. This procedure involved viewing a segment and then immediately editing the paragraph which discussed the segment. When the entire test was completed, the draft answer keys were shared with the panel of experts and discussed along with the test paragraphs to obtain recommendations for revision.

10. Score panel of expert edited response sheets; tally the results to determine amount of agreement: The panel of expert response sheets were scored in the same manner scoring of student responses was anticipated to occur. This provided a test of the scoring procedure and materials as well as the paragraph content. A scoring template was prepared for each paragraph. The template provided a quick, hand scoring approach since the development of computerized scoring was not possible within project resources. The template was a transparency made from an answer key (see step 8 above) in which the paragraph content was deleted and the item numbers were retained. Colored lines were drawn on the transparency underlining items which were to be deleted in one color and those to be retained in another color. This template was then simply superimposed on an individual's response sheet and a score entered for each item. A 1 (one) was entered for an item correctly crossed out or left in and a 0 (zero) was entered for an item incorrectly crossed out or left in.

Video segments eliciting a wide disparity in the way members of the panel of experts edited the accompanying test paragraph were deleted. Those segments eliciting total or almost total agreement were retained. Segments eliciting a

moderate amount of agreement were retained and the paragraphs revised to reflect panel of expert recommendations. A TRT test with 12 test video segments and 2 practice segments resulted from this process and constituted the material that was pilot tested in step 11. The number of items in each test segment ranged from 18 to 45. The total number of items across all twelve test segments was 330.

Readability of the test segment paragraphs was analyzed and the results used in revising the paragraphs. Readability estimates of the revised paragraphs are presented in Table 1 below. A high school level of readability was sought in order to minimize the contribution of reading capability to the obtained score on the paragraphs. However, because of the length of the sentences required in order to preserve serial listing of items in a single sentence, readability of some paragraphs remained at a collegiate level. Further, some terminology in the paragraphs was not common terminology and influenced the readability level of the paragraphs in an upward direction. Because this terminology would be familiar to students studying the professional area, this factor was thought to inflate the reading level estimates for this population.

Table 1

Readability of TRT Paragraphs

	Fry Method	Flesch Method	
Segment	Grade Level	Style	Grade Level
56	8	Standard	8-9
99	8	Standard	8-9
26	7	Standard	8-9
27	12-college	Difficult	13-16
29	10-12	Fairly Difficult	10-12
32	12	Difficult	13-16
05	10	Fairly Difficult	10-12
34	college	Difficult	13-16
12	11	Fairly Difficult	10-12
07	college	Difficult	13-16
53	college	Difficult	13-16
82	college	Difficult	13-16
69	11-12	Fairly Difficult	10-12
90	9	Fairly Difficult	10-12

11. A video tape was produced containing the TRT test segments resulting from steps five through ten. The segments were sequenced on the tape in the following order: the two practice segments; shorter, less complex segments; and longer, more complex test segments. The TRT test was pilot tested as described in the pilot testing procedures section below.

12. The final step in the process of TRT test development was review of the pilot test data and the test paragraphs and scoring keys by the panel of experts. Following this review, three test segments with the lowest internal consistency reliability coefficients were eliminated, leaving nine test segments. No revision of response sheets and answer keys was done on the remaining test segments in light of strong internal consistency indicated for these segments by the pilot test data (see Chapter 3).

TRT Pilot Testing Procedures

Test booklets were assembled. Each booklet contained a cover sheet for recording segment scores and coding of demographic information. An instruction sheet followed by a demographic questionnaire appeared next. These were followed by sections relating to each of the fourteen video segments in the test. Each section pertaining to a video segment contained a cover sheet which indicated the title of the segment and a one or two sentence description of the segment for the purpose of orienting the student to the segment. A seal prevented the student from viewing the rest of the section until after they had seen the video segment. The rest of the section contained the test paragraph and a grid for recording the score on each item contained in the paragraph. The sections which accompanied the two practice segments contained one additional item. This item was a copy of the test paragraph which had been edited to show the parts to be crossed out and the parts to be retained.

The TRT test was administered to a sample of 123 students and 4 instructors. Students included: 19 high school students enrolled in a child development course; 47 technical institute students enrolled in a one-year child care program; 22 community college two-year, associate degree early education students; and 35 baccalaureate degree and post-baccalaureate college and university students.

Administration of the TRT took approximately two hours. The purpose of the pilot testing was shared with the students followed by the presentation of the instructions for taking the test. Students were asked to read the instructions in their test booklet as an audio tape of the same instructions was

played. Students were given an opportunity to ask questions about test procedures at the end of the presentation of instructions, and before the practice portion of the test was initiated.

The practice portion of the test involved asking students to read the title and description of a video segment, showing the video segment, telling the students to break the seal on that section of their test booklet and to edit the test paragraph. Students were asked to work on the demographic questionnaire in their test booklet if they finished their editing before others were done. Students were given as much time as they needed to finish editing the paragraph. When all students had completed the editing task, they were asked to break a second seal, which only the practice segment sections contained, to reveal the correctly edited paragraph. The practice portion contained two video segments, giving the students an opportunity to experience the full testing procedure twice. At the end of the practice portion of the test, students were again given an opportunity to ask questions about test procedures.

Once the two practice segments had been completed, the twelve test segments were presented in a similar manner. Students were reminded at the beginning of this portion of the test that the paragraphs they edited from now on would be scored and to do their best work.

Test Scoring Procedures

The cover sheet of each completed test booklet was given an identification number and a group code. The group code signified which test sample the student was in and enabled categorization of the student's data in the appropriate educational level. The identification number was then transferred to each page of the test booklet.

The test booklets were then separated and the edited paragraphs of all respondents for each segment were grouped together for scoring. Scoring involved laying the scoring template over each edited paragraph and recording a 1 (one) in the scoring grid by each item the student had correctly crossed out or left in and a 0 (zero) for each item the student had incorrectly crossed out or left in. A segment score was calculated for each segment for each student by adding the number of 1s in the scoring grid.

Data Analysis

Item and segment scores were analyzed to determine internal consistency reliability of the test using the KR-20 statistic. Segment scores were correlated using the Pearson r statistic as a measure of internal test validity. Total scores, obtained by

adding unweighted segment scores, and demographic data were analyzed using Pearson r correlations and mean comparisons to reveal external test validity.

The instructor data was analyzed with the baccalaureate degree and post-baccalaureate college and university students data rather than as a separate group because of the small number of instructors and because the educational background of the instructors was similar to this group. Results of the pilot test and analyses to evaluate the TRT are reported in Chapter 3.

CHAPTER 3

PILOT TEST RESULTS, SUMMARY, CONCLUSIONS, IMPLICATIONS AND RECOMMENDATIONS

This chapter reports the results of the pilot test of the TRT. Pilot test data were analyzed to determine the reliability and validity of the TRT. Experience with the TRT during the pilot test was used to evaluate the TRT's practicality.

Reliability

Reliability of the TRT was determined by dividing the 127 scores on each segment into two groups, a high scoring group and a low scoring group, and computing a KR20 coefficient for each segment. The KR20 statistic provides an indication of internal consistency reliability for test items scored 1, 0, correct or incorrect. The KR20 coefficients reported in Table 2 indicate the internal consistency for items within each segment of the TRT.

Table 2

Internal Consistency Reliability of TRT Paragraphs (N=127)

Segment No.	No. of Items	Variance	KR20 Coefficient
26	(18 items)	6.396	.704
27	(28 items)	20.982	.833
29	(31 items)	23.540	.847
32	(24 items)	21.256	.826
05	(27 items)	30.609	.887
34	(20 items)	24.855	.877
12	(22 items)	10.641	.697
07	(25 items)	16.361	.829
53	(26 items)	7.016	.671
82	(28 items)	9.021	.668
69	(36 items)	21.143	.766
90	(45 items)	76.438	.910

The three segments with the lowest reliability, segments 12, 53 and 82, were eliminated from the final version of the test based on these and other pilot test data. The internal

consistency reliability coefficients for each of the remaining segments are well within the acceptable range for a number of uses including classroom testing, diagnosis, and program evaluation. Further, the strong reliabilities indicate that the TRT test may prove useful for personnel selection decisions if further research is conducted that would justify its use for such purposes. Because the test items were embedded within paragraphs which related to specific and different situations, no internal consistency reliability tests were performed for all the test items in relation to each other.

Internal Validity

Internal validity of the TRT was investigated by analyzing the degree of association among the test segments. Each segment in the test is scored separately and is assumed to make a unique contribution in measuring a different aspect of the student's functioning with respect to the content/practice domain. It would be expected that intercorrelations among test segments would be low to moderate and that high intercorrelations should not occur if segments are truly tapping different and unique aspects of expertise. Low intercorrelations indicate a high degree of independence among test segments. That is, a student might do very well on one segment and very poorly on another. Such a pattern would indicate that each segment is contributing in a highly unique way to assessing the student's functioning within the content/practice domain. Table 3 presents the intercorrelation matrix that indicates the correlation (Pearson r coefficient) of every test segment with every other test segment. A coefficient of 0.0 indicates no relationship; a coefficient of 1.0 indicates a perfect positive relationship; a negative coefficient indicates a relationship exists in opposing directions. The number of scores varies slightly for some of the segments since a few students did not complete all segments.

Correlations of each segment score with the total score (obtained by adding unweighted segment scores) are presented in Table 4. Total scores were obtained only for those students who completed all segments.

The low to moderate intercorrelations between segment scores presented in Table 3 indicate that segments operated highly to moderately independently from one another. The moderate correlations (the highest of these was .49) between segment scores indicate that the segments either measured aspects of the content/practice domain that have some association, or that there is overlap among the segments on the same factor(s). The moderate correlations do not suggest that there is sufficient overlap between any of the segments to conclude that any two of them tap largely the same aspects of the content/practice domain. The degree of independence apparent among the segments would suggest that caution should be exercised in interpretations of

Table 3

Pearson r Correlations Indicating Degree of Relationship Between Test Segments (N=113-127)

Segment Number	Segment Number											
	26	27	29	32	05	34	12	07	53	82	69	90
26	1.00	.09	.27	.27	.18	.03	-.02	.05	.06	.11	.07	.19
27		1.00	.19	.27	.26	.01	-.11	.15	.13	.09	.13	.22
29			1.00	.49	.44	.06	-.09	.20	.12	.08	.18	.42
32				1.00	.33	.03	-.11	.07	.11	.12	.07	.33
05					1.00	.23	-.15	.09	.06	.14	.18	.22
34						1.00	.18	.01	.22	.23	-.14	-.17
12							1.00	-.05	-.00	.07	-.21	-.18
07								1.00	.33	.15	.22	.06
53									1.00	.15	.20	.22
82										1.00	.03	.06
69											1.00	.24
90												1.00

Table 4

Relationship Between Segment Scores and Total Score (N=113)

Segment No.	Correlation (r) With Total Score
26	.37
27	.46
29	.69
32	.59
05	.61
34	.27
12	-.02
07	.38
53	.43
82	.35
69	.39
90	.63

relationships between the unweighted total scores and other variables. Clearly, more research is needed before it can be determined if and how segment scores can be combined. For purposes of this report, total scores should be viewed as an

extremely rough indicator of overall performance on the test. The correlations of the segments with the total score (Table 4) indicate the degree to which performance on each segment was consistent with the overall test performance. The moderate correlations obtained for some segments reflect the relative independence of these test segments. The higher correlation of other segments (e.g., .69, .63, and .61) indicates a tendency for performance on these segments to be more closely related to performance on the test overall. The one exception to this pattern is segment 12. Segment 12 reflects a different pattern from the other segments, both in the negative correlations between segment scores indicated in Table 3 and the lack of association with the total scores indicated in Table 4. This segment seemed to be operating in a different manner from all of the other segments and may be of questionable validity. For this reason, and because of its lower internal consistency, this segment was eliminated from the final version of the test.

External Validity

Several analyses were conducted to provide information regarding the validity of the experimental TRT with regard to variables external to the test. These variables were educational level, grade point average, and experience with children. Data regarding these variables were obtained via the demographic questionnaires that were included in the test booklet and, in the case of grade point averages, from the schools in which the students were enrolled. It was expected that, if the TRT is successful in eliminating surface cues to the right answer and is a true reflection of the individual's knowledge and expertise in a content/practice domain, the correlations with these external variables would be moderate. This was because, first, it is assumed that grade point averages reflect to some extent a student's general intelligence and ability to take tests irregardless of the extent of their knowledge and capability in an area. Second, it is assumed that experience provides a certain level of knowledge and capability but is not sufficient by itself for developing a high level of expertise because not everyone's practice improves beyond a certain level as a result of experience. Third, it was assumed that the more education an individual had relevant to the content/practice domain, the higher their scores would be since education would be expected to affect knowledge and expertise in a content/knowledge domain to a point. However, like experience, it was expected that education would be influential only to a degree and that a high level of expertise requires more than education. In other words, education and experience contribute to but do not guarantee, expertise. Table 5 reports the results of the analyses in which Pearson r was used to determine the degree of association of the total test scores with these external variables. Again, caution is warranted in interpreting these correlations based on total

Table 5

Relationship of Total Score and External Variables

Variable	N	r
Educational level	122	.481
Overall grade point	56	.455
Grade point in major	52	.472
Number of children parented	21	.243
Number of months as an aide working with children in day care, nursery school, elementary school, or a family day care provider	87	.098
Number of months as a teacher in day care, nursery school, or elementary school	17	.198

scores given the need for further investigation of the function and formation of total scores.

The moderate correlations with most of the external variables were at anticipated levels and in directions that support the validity of the TRT. Educational level and grade point in major had the strongest relationship with the TRT total scores. This provides evidence of the important, but insufficient role of knowledge in level of expertise. The slightly stronger correlation of total score with major field grade point average is in a direction that supports the claim that the TRT measures knowledge application in a specific knowledge domain. However, too much stock should not be placed on this slight difference since the correlation between overall and major grade point average was .96. Care should be observed in interpreting the external validity coefficients since the number of cases varied considerably for each variable. Grade point data was not accessible for all the students who participated in the study.

The correlations for parenting and work experience with children are suggestive of external validity of the TRT. They suggest that amount of experience with children at higher levels of responsibility may be related to level of total score whereas amount of experience with children at lower levels of

responsibility is not. Since level of responsibility is usually associated with educational level (although this would not hold true for number of children parented), it is likely that the correlations for experience reflect educational level to some extent. It should also be noted that the numbers of individuals with higher level responsibility experience was low, making correlations only suggestive.

A second way in which educational level was analyzed with respect the performance on the TRT was to examine the pattern of median and mean scores on the TRT across the educational levels included in the pilot test. Table 6 reports this comparison along with other descriptive statistics regarding total scores.

Table 6

Comparison of Total Scores by Educational Level (N=113)

Current Educational Level	N	Range	Median	Mean	S D
High school	14	206-262	236	233	17.4
1 year post-secondary program	45	185-279	236	236	21.4
2 year post-secondary program	22	215-298	254	257	21.4
Baccalaureate or post-baccalaureate program	32	226-305	267	267	20.2

The median and mean total scores increased by educational level in an expected pattern, with the high school group showing the lowest median and mean scores and the baccalaureate and post-baccalaureate group showing the highest median and mean scores. Post-baccalaureate students were not analyzed as a separate group since several of these students were enrolled in licensure rather than degree programs. Their backgrounds in child development and early education varied considerably. For some

of these post-baccalaureate students, background in child development and early education was not dissimilar to that of baccalaureate program child development and early education students even though they held a degree in a related or unrelated field.

The greatest range in total scores was in the one year postsecondary program group. These were the technical institute day students combined with a few evening program students. The broad score range in this group likely reflects the presence of a considerable proportion of students with learning disabilities as indicated by their instructors. The degree to which the dependence on reading in taking this test was a factor in these scores needs to be considered in interpreting these results. Other than this exception, the high school students had the lowest low and high scores. The baccalaureate and post-baccalaureate program students had the highest low and high scores, patterns that support the external validity of the TRT. The two year program students' high score and median score fell between the high and median scores of the one year program students and the baccalaureate and post-baccalaureate program students. The low score and median score for the two year program students fell between the same two groups as expected. The one year program students' high score fell between that for the high school and two year program students' high score as expected. The median score for this group was the same as that for the high school group, pulled down by the low scores which some individuals in this group attained as explained above.

To summarize the validity data, the patterns in the range, median and mean data are supportive of the external validity of the TRT. Relative to the external variables of educational level and grade point average, the TRT performed as expected for each variable, showing moderate correlations with each external variable and a steadily increasing mean and median score as educational level increased. Only one exception in the data pattern exists. One year program students scored lower and also higher than high school students. The unexpected lower scores are likely attributable to a high proportion of students with learning disabilities in this group.

Usability

Usability of the TRT will be discussed with respect to practical issues such as time, equipment, materials and expertise requirements.

The TRT that was pilot tested took approximately two hours to administer, allowing students as much time as they needed to complete the editing of the paragraphs. The final version of the test which has three less segments than the pilot tested version takes less time. This time frame fits well within

reason for use in classroom testing situations. Because the segments are relatively independent, splitting the administration into two or even three sessions is acceptable. However, it would be important to do one of the warm-up/practice segments at the beginning of each session subsequent to the first session when both practice segments should be used.

Students learned the editing process very quickly and their speed in editing the paragraphs increased noticeably after the second test segment (i.e., after completing four segments, two practice segments and two test segments).

Equipment requirements for the test include a video tape player and a monitor. Needed materials include test booklets containing the test paragraphs, a video tape containing the test scenes, soft lead pencils (No. 2s), and scoring templates for each segment. TRTs that do not use video tape as stimulus material would not need a video tape player and monitor but may need other equipment required by a different type of stimulus material.

Expertise requirements for test administration are well within the resources schools have. The test can be administered by classroom teachers or evaluators. The test does not need to be administered by someone with specialized expertise in the content/practice domain or in testing. Specialized expertise is not needed for scoring the test. Scoring takes approximately one minute per segment per student, making the total scoring time required on the nine segments remaining in the final, revised version of the test nine minutes per student.

Given these requirements, it would seem that the TRT, as a testing approach, could be used in any vocational education field and classroom or laboratory and in any subject within secondary or postsecondary curricula where a test of thinking with and application of knowledge within a content or practice domain is desirable.

It became apparent throughout both the development and the pilot testing process that the TRT is likely to be useful for instructional purposes as well as for assessment purposes. Sharing students' scored, edited paragraphs with them is likely to provide feedback that has a high degree of instructional value, especially if the test is gone over by replaying the scenes and discussing discrepancies between the ways students edited the paragraphs and the paragraph scoring key. The video taped scenes with or without the test paragraphs provide rich material for discussion, student hypothesis generation and problem solving as part of a learning process.

Summary, Conclusions and Implications

The pilot test findings were interpreted in relation to the purposes and objectives of the study and the criteria against which the TRT was to be evaluated that were indicated in Chapter 1. The purpose of the study was to investigate the potential of the Tailored Response Test (TRT) as an approach to assessing higher order thinking and knowledge use within a specific knowledge domain and area of practice. Specifically, the TRT was investigated for its potential to:

- assess student learning of perceptual and thinking processes and complex judgment processes relevant to work roles
- go beyond assessing possession of knowledge to assessing students' ability to use their knowledge

Objectives of the study were to develop and evaluate a prototype TRT as an approach to assessing learning within a specific knowledge domain and practice area relevant to vocational education.

A TRT was developed in a content/practice area relevant to vocational education. The TRT was tested in child development and care programs funded by vocational education monies and in early education programs. The TRT appears to tap student learning of perceptual and thinking processes and complex judgment processes relevant to work roles. This is reflected in the external validity data which indicated moderate correlations between total test scores and educational level, and grade point average. The moderate correlations with grade point suggest that something beside knowledge storage is reflected in the test scores, given that grade point to a large extent is the result of objective tests that measure knowledge storage and recall. The test was designed to require judgment processes based on information obtainable through strategic perception and thinking and appears to have accomplished this objective. The data support the conclusion that the TRT that was developed does go beyond assessing possession of knowledge to assessing students' ability to use their knowledge.

How well did the TRT fare with respect to the criteria against which it was to be evaluated as indicated in Chapter 1?

Criterion 1: Focus on the totality of situations rather than on only one or two aspects. The video-taped scenes captured the totality and complexity of situations with high fidelity and in a concentrated fashion. Very short episodes contained many factors, presenting them with the subtlety reflected in real life events. However, only video-taped scenes which had

excellent clarity in audio and visual quality met this criterion fully. Voice tone, glances, and subtle body motions all provided cues important to understanding the situation and had to be discernible to accurately understand and comprehensively evaluate the situation.

Criterion 2: Focus on higher level thinking processes. Since many of the paragraphs focused on what was appropriate, evaluative judgment was required to edit them. Editing the paragraphs required a combination of interpretation and judgment processes. More research is needed to clarify the exact nature of the thinking required to successfully edit a TRT paragraph. However, the present research study suggests that more than memorized knowledge is operating in scores (suggested especially in the grade point-total score correlations). Because of the way the paragraphs were constructed and their focus on evaluative judgment, it is likely that this "something more" is, in large part, higher level thinking.

Criterion 3: Responses reflect expertise in an identified body of knowledge. The external validity data which indicates increasing mean and median total scores with increasing educational level suggests that responses do, indeed, reflect expertise in an identified body of knowledge. This conclusion could be further confirmed by research that compares performance on the TRT of an expert group and one with no preparation or experience in the area.

Criterion 4: Can be administered within the time structure of typical classrooms. This criterion was met even with the pilot test which contained three more segments than the final version of the test contains. The independence of the segments which makes testing across more than one session more feasible is an added advantage in fitting the time structure of typical classrooms.

Criterion 5: Easily scorable in a short amount of time. This criterion was met. Scoring of the final version of the test should not take more than nine minutes per student using a scoring template as in the pilot test. This means that six or seven tests can be scored per hour. Computerized scoring can be developed which would cut scoring time dramatically. Computerized scoring requires that the paragraphs be printed on light sensitive paper, that a scanner be available, and that a computer program be written for the scoring procedure. Scoring with a template could be done by a classroom teacher.

Criterion 6: Assumes that real-life problems don't have one right answer but rather, have several possible resolutions which may be appropriate and none of which may be perfect. This criterion was partially met. The assumption that real-life problems don't have one right answer but, instead, have several

possible resolutions, is reflected in the way the paragraphs are constructed to emphasize whether or not an action is appropriate. The language in the paragraphs does not speak to right and wrong but rather to more and less appropriate or to appropriate and inappropriate. However, there is still a "right way" to edit the paragraphs. This is a limitation that the desire and practical need for quick scoring procedures produce. A further limitation is that test respondents do not have the opportunity to record ideas they generate. For example, a student might have a particularly creative thought about how one of the situations might have been handled. There is no opportunity in the TRT for them to record that thought or get credit for it. Perhaps further research and adaptation regarding the TRT will reduce this limitation. However, it seems apparent that to the extent that quick scoring is a high priority in testing, the possibility for student generated responses is reduced.

Criterion 7: Avoid potential for test score to reflect "test-wiseness" - the ability to guess the right answer from surface characteristics of the items. This criterion was met. No indication of where the items are is present in the paragraph. Further, there was a deliberate attempt in constructing the paragraphs to avoid using a particular type of sentence structure for items to be deleted or for items to be retained. In addition, paragraphs were carefully constructed so that parts of sentences were not interdependent so that if one part was crossed out one would automatically know if other parts should be deleted.

Criterion 8: Is reliable and valid at levels acceptable for classroom use, program evaluation, and for potential standardization. The reliability criterion was met for internal consistency within test segments. Segment internal consistency reliabilities were well within a range acceptable for classroom use. The large proportion of segments with internal consistency reliability coefficients in the high 80s and low 90s suggest that the test has potential for standardization. Further research is needed to pursue questions raised by internal and external validity data presented here. Support for external validity was suggested but should be pursued further since, as with any test, validity is established over time and through data obtained as the test is used in various situations and with various other measures. Internal validity questions, particularly those which stem from the low to moderate correlations among segment scores, must be investigated before sound conclusions can be reached and before a total score should be used to represent a student's performance on the test. While the investigation of test validity should not end, the pilot test evidence provides sufficient support for the test's validity to justify its use for classroom testing and program evaluation purposes, at least on an experimental basis.

The tailored response test developed in this project provides several useful items. First, it provides a test that measures learning that is sophisticated and difficult to measure - the integration of knowledge with thinking processes in the care and guidance of children. Such learning has not been satisfactorily measurable in the past and, as a result, the ability of programs to document in a reliable and valid manner achievement of significant educational program goals has been limited. Second, the experimental TRT developed in this project provides a prototype of a test that measures the integration of knowledge with thinking processes that is usable in a wide variety of fields of study - fields within vocational education and beyond it. Because of the great flexibility in the types of stimulus material that can be used in the TRT (video tape, audio tape, charts, graphs, text, etc.) the TRT is likely to find use in a wide range of areas. Third, the TRT provides an alternative approach to objective testing and one that is likely to spawn further adaptations and creative thinking about the possibilities for new types of testing materials that are better suited to and more capable of measuring the mental structures and processes underlying expertise than are present types of tests.

Recommendations

Although further research on the TRT developed in this project is needed to answer various deeper questions about it, it is usable now in educational programs concerned with preparing students and personnel for work with young children. It should be used extensively to gain experience with it, to develop a base of further data about it, and to address the needs such educational programs have for such an assessment device.

Further research should be done that investigates reading levels of the paragraphs in relation to segment score levels of different types of students. Stability reliability should be established for the test. The investigation of test validity, both internal and external, should continue. Research that analyzes in detail and depth specific mental processes involved in editing paragraphs is needed. External validity data which indicates the degree of association between TRT scores and actual judgment making by the test taker in real situations where they are responsible for children is needed. Ways of adapting the TRT should be investigated. Ways of combining paragraph editing with student generated material should be pursued. Another potential adaptation that should be investigated would involve sharing additional information about the stimulus material situation after the student has seen the situation and edited the paragraph, and asking the student to edit a different copy of the same paragraph again to reveal the student's ability to adapt their assessment of a situation based on new information. The TRT should be experimented with as a diagnostic tool especially

since it seems to be able to reflect mental processes and knowledge structures over a wide continuum. Its use as a teaching tool should also be explored. Finally, an interactive video disk, computerized version of the test should be explored.

In summary, the TRT developed here provides an assessment tool that is "ready for use" by teachers in child development, child care and early education programs, and that also has instructional potential. The TRT also provides a starting place for the development of additional TRTs in other educational areas. It would be desirable for teachers to be involved in the development of TRTs. Since the process is well defined, this should be a possibility, especially if school district or other evaluation personnel are available to assist with pilot testing design and data analysis.

REFERENCES

Baldauf, R. & Propst, I. (1979). Matching and multiple-choice Cloze tests. Journal of Educational Research, 72(6), 321-326.

Compton, N. & Hall, O. (1972). Foundations of home economics research: A human ecology approach. Minneapolis: Burgess.

Educational Testing Service. (1987). Assessment in the service of learning. Proceedings of the 48th ETS Invitational Conference. Princeton, NJ: Educational Testing Service.

Flanagan, J. (1954). The critical incident technique. Psychological Bulletin, 51, 377-358.

Frederiksen, N., Saunders, D., & Wand, B. (1957). The in-basket test. Psychological Monographs: General and Applied, 71(9, Whole No. 438), 1-28.

Glaser, R. (1984). Education and thinking: The role of knowledge. American Psychologist, 39, 93-104.

Good, C. (1972). Essentials of educational research: Methodology and design. Englewood Cliffs, NJ: Prentice-Hall.

Haney, W. V. (1985). Making testing more educational. Educational Leadership, 43 (2), 4-13.

Jonz, J. (1976). Improving on the basic egg: The M-C Cloze. Language Learning, 26(2), 255-265.

Kostelnik, M., Stein, L., Whiren, A. & Soderman, A. (1988). Guiding children's social development. Cincinnati: South-Western.

Larkin, J. (1979). Information processing models and science instruction. In J. Lockhead & J. Clement, (Eds.), Cognitive process instruction (pp. 109-118). Philadelphia: The Franklin Institute Press.

Lave, J. (1985). Introduction: Situationally specific practice. Anthropology and Education Quarterly, 16, 171-176.

Manning, W. (1985). Development of Cloze-edit tests of English as a second language. Unpublished manuscript. Educational Testing Service, Princeton, NJ.

Marshall, S. (1988, April). Assessing schema knowledge. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

McKenna, M. (1986). Cloze procedure as a memory-search process. Journal of Educational Psychology, 78(6), 433-440.

Miles, M. & Huberman, A. (1984). Qualitative data analysis. Beverly Hills: Sage.

Quellmalz, E. S. (1985). Needed: Better methods for testing higher order thinking skills. Educational Leadership, 43 (2), 29-35.

Taylor, W. (1953). Cloze procedures: A new tool for measuring readability. Journalism Quarterly, 30, 415-453.

Taylor, W. (1956). Recent developments in the use of the cloze procedure. Journalism Quarterly, 30, 415-433.

Thomas, R. & Litowitz, L. (1986). Vocational education and higher order thinking skills: An agenda for inquiry. St. Paul: University of Minnesota, Minnesota Research and Development Center for Vocational Education.

Thomas, R. (Ed.). (1988). Thinking underlying expertise in specific knowledge domains: Implications for vocational education. St. Paul: University of Minnesota, Minnesota Research and Development Center for Vocational Education.

Thorndike, R. (1949). Personnel selection: Test and measurement techniques. New York: John Wiley & Sons.

Thornton, R. (1988). Evaluation of a scannable method of scoring open-ended-response simulations. Unpublished manuscript, Educational Testing Service, Princeton, NJ.

Willson, V., Kulikowich, J., Alexander, P. and Farrell, D. (1988, April). A cognitive theory for test design: Implications for assessing domain-specific and strategic knowledge. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Witkin, B. (1984). Assessing needs in educational and social programs. San Francisco: Jossey-Bass.

APPENDIX

Content/Practice Domain for

Tailored Response Test

Content/Practice Domain

Supporting Children's Social Development

Function: I. By Building Positive Relationships Through Nonverbal Communication

Element: A. By the adult using positive, nonverbal communication with the children

Tasks:

1. the adult observes the nonverbal behavior of the children in her/his care
2. the adult recognizes the cultural and family variations in children's nonverbal behavior
3. the adult maintains the integrity of the children's proximal space
4. the adult uses nonverbal signals to gain the attention of a group of children who are engaged in an activity or who are dispersed in space
5. the adult walks up to the children with whom she/he wants to communicate and orients her/himself in a face-to-face position
6. the adult keeps all channels of communication consistent when communicating about her/his feelings
7. the adult touches the child
8. the adult stands, sits, or squats close to the child, not more than an arm's length away
9. the adult sits or stands so that her/his head is at the same level as the child's head
10. the adult maintains frequent but not continuous eye contact with the child
11. the adult faces the child so that her/his shoulders and the child's shoulders are parallel
12. the adult leans slightly toward the child
13. the adult participates in nontraditional gender role activities

Element: B. By the adult showing warmth and caring

Tasks:

1. the adult conveys a generally positive facial expression in neutral situations
2. the adult responds as quickly as possible when spoken to, and takes the time to listen
3. the adult uses voice tones that are normal to soft in loudness and normal to low in pitch and a voice quality that is relaxed, serious, and concerned when talking with the children

Element: C. By the adult demonstrating acceptance and security

Tasks:

1. the adult maintains a tone of voice that is firm, warm, and confident--the pitch should be even and the normal volume
2. the adult relaxes, maintains close physical proximity, and maintains arms and legs in an open or semi-open position
3. the adult uses her/his hands to gesture appropriately or if necessary, to grasp the child until the communication is complete
4. the adult is accepting of both sexes participating in nontraditional gender role activities

Function: II. By Promoting Children's Self-Awareness And Self-Esteem

Element: A. By the adult formulating the fundamental skills associated with a positive verbal environment

Tasks:

1. the adult greets the children when they arrive
2. the adult addresses the children by name
3. the adult extends invitations to the children to interact with her/him
4. the adult speaks politely to the children
5. the adult listens attentively to what the children have to say
6. the adult invites the children to elaborate on what they are saying

7. the adult thinks of some conversation openers in advance
8. the adult remains silent long enough for the children to gather their thoughts
9. the adult takes advantage of spontaneous opportunities to converse with the children
10. the adult refrains from speaking when talk would destroy the mood of the interaction
11. the adult uses encouraging responses so the children feel better about themselves
12. the adult uses positive verbal reinforcement for cross gender role activities and play
13. the adult helps the children express pride in their own cultural heritage

Element: B. By the adult formulating behavior reflections

Tasks:

1. the adult describes some aspect of the child's person or behavior in a statement to the child
2. the adult phrases behavior reflections as statements
3. the adult addresses behavior reflections directly to the children

Element: C. By the adult formulating paraphrase reflections of the children's verbal expressions

Tasks:

1. the adult listens actively to the children's words
2. the adult restates in her/his own words what the child has said
3. the adult rephrases her/his erroneous reflections
4. the adult matches her/his reflections to each child's ability to understand language
5. the adult uses a conversational tone when reflecting
6. the adult summarizes the children's actions and words rather than reflect each individual behavior or idea expressed
7. the adult selects one idea at a time to paraphrase from the many the child may express

8. the adult adds interest to her/his reflections by periodically phrasing them in a form opposite from that used by the child
9. the adult reflects first when the children ask her/him a question

Element: D. By the adult formulating questions

Tasks:
1. the adult asks open-ended questions
2. the adult asks questions when she/he is truly perplexed
3. the adult carefully chooses open-ended questions
4. the adult emphasizes quality over quantity in using questions in conversation with children

Element: E. By the adult promoting self-esteem in all children

Tasks:
1. the adult makes every child the object of daily focused attention
2. the adult treats all of the children with respect regardless of gender, ethnic background, social class, and/or handicapping condition

Function: III. By Responding To Children's Emotions

Element: A. By the adult formulating affective reflections of the children's emotions

Tasks:
1. the adult observes the children carefully before saying anything
2. the adult is sensitive to the wide range of emotions that the children exhibit
3. the adult makes a nonjudgmental assessment of what the child is experiencing
4. the adult makes a brief statement to the child describing the emotion she/he observed
5. the adult uses a variety of feeling words over time
6. the adult acknowledges the children's emotions even when she/he does not approve of them
7. the adult revises inaccurate reflections

Element: B. By the adult helping the children to increase their verbal expressions of emotions

Tasks:

1. the adult sets an example for talking about emotions by bringing them up her/himself
2. the adult explains to children who are involved in emotional situations that they can tell their emotions to the other person
3. the adult assists children in describing their emotions to others if they cannot do so entirely on their own
4. the adult helps children decipher behavioral cues that tell how another person is feeling
5. the adult draws the children's attention to situational cues that contribute to people's emotions
6. the adult expands the children's vocabulary to facilitate communication of troubled feelings and thoughts
7. the adult encourages both boys and girls to express their emotions
8. the adult increases the children's sensitivity to their own body sensations when they feel angry, sad, tense, joyful, and so on
9. the adult teaches the children to practice positive self-talk in tense situations

Element: C. By the adult helping the children to cope with difficult emotions

Tasks:

1. the adult acknowledges the children's negative emotions and forbids destructive actions
2. the adult comforts children of both sexes who are sad or afraid
3. the adult helps children sort out mixed emotions
4. the adult provides children with information that may enlarge their perception of a situation
5. the adult provides opportunities for children to observe how others of both sexes cope in a situation they fear
6. the adult allows children to approach a feared situation gradually

7. the adult helps children think of new strategies or learn new skills as a way to deal with difficult emotions
8. the adult gives the children opportunities to work out their feelings through play
9. the adult uses ordinary experiences and daily activities to discuss feelings, thoughts, and behaviors that people use when they are afraid, uncertain, faced with change, or overwhelmed by what is happening to them
10. the adult anticipates and rectifies situations in which children may be unduly frustrated
11. the adult remains alert to children for whom frustration is building

Function: IV. By Enhancing Children's Play

Element: A. By the adult setting the stage for the children's play

Tasks:

1. the adult establishes the necessary conditions for play
2. the adult says to her/himself, "It's okay to play, to laugh, to have fun."
3. the adult stands or sits near children at play
4. the adult pays attention to what the children are playing, and what they say and do
5. the adult schedules playtime in segments that are long enough for play concepts to be developed
6. the adult prepares the children for a change in activities by warning them in advance that the change will occur
7. the adult sends the children to an activity rather than away from one
8. the adult provides adequate space for the number of children playing
9. the adult provides quality playthings for all types of play for both boys and girls
10. the adult remains alert for valuable learning experiences that may be created spontaneously by the children, being flexible enough to let them progress without interruption
11. the adult rechannels group play in which children pretend to kill one another

Element: B. By the adult maximizing the play potential of the materials available

Tasks:

1. the adult mixes unrelated toys together
2. the adult introduces novel toys and materials slowly
3. the adult rotates playthings
4. the adult arranges the materials to encourage interaction between children
5. the adult encourages the children's novel use of more traditional (gender related) toys

Element: C. By the adult helping children to acquire skills through the direct involvement of an adult as a player

Tasks:

1. the adult plays with the materials
2. the adult takes a role to encourage pretend play
3. the adult demonstrates movements as necessary
4. the adult participates fully in the game
5. the adult demonstrates that she/he can play in nontraditional play roles (cross gender)

Element: D. By the adult helping individual children change the level of their social participation in play

Tasks:

1. the adult observes the child for cues that the present level of participation is inadequate
2. the adult matches the activity to the child's level of skill
3. the adult plays with the child
4. the adult invites the child and a second player to play with her/him, then eases out of the situation

Element: E. By the adult escalating the level of play gradually by varying her/his play performance or by giving cues through play signals or metacommunications

Tasks:

1. the adult extends object play by imitating what the child is doing, then varies the activity a little

2. the adult suggests that children use specific play signals to initiate or sustain play
3. the adult withdraws from the play and resumes the role of observer once the play is well under way

Element: F. By the adult coaching children occasionally from outside the play frame

Tasks:

1. the adult suggests a related theme
2. the adult adds a necessary prop
3. the adult introduces new players from outside the play frame
4. the adult teaches players to use a clear signal when leaving the play frame
5. the adult makes suggestions to further the goals of the children, such as pointing out a problem or restating game rules
6. the adult teaches the children games when necessary
7. the adult encourages the children when they play with nontraditional (gender linked) toys and play in nontraditional (gender related) roles

Element: G. By the adult becoming directly involved in the children's playfulness

Tasks:

1. the adult demonstrates a nonliteral approach to resources
2. the adult is accepting of young children's humor
3. the adult explains that a child was only joking when someone misinterprets the meaning of what was said or did not recognize a play signal
4. the adult uses affective reflections when children laugh at disfigurement, falls, or handicapping conditions; then, provides brief but accurate information

Element: H. By the adult demonstrating her/his awareness of children's individual differences

Tasks:

1. the adult accepts the young child's approach to games with rules

2. the adult matches the play activity with the skills of the players regardless of gender
3. the adult accepts the child's play style preferences regardless of gender
4. the adult provides support for children when other children out perform them in play
5. the adult supports the children in their choice of play activities; not limiting play to sex-stereotyped choices

Function: V. By Fostering Self-Direction In Children

Element: A. By the adult reflecting on problem situations

Tasks:

1. the adult observes the children carefully before reflecting
2. the adult formulates reflections that accurately describe the child's perspective
3. the adult reminds her/himself to describe the child's point of view before her/his own
4. the adult pays attention to children's age when deciding which type of reflection to use
5. the adult avoids using "but" as a way to connect the reflection to the rest of the personal message
6. The adult mediates the children's conflicts
   a. the adult initiates the mediation process
   b. the adult clarifies each child's perspective
   c. the adult sums up the situation
   d. the adult assists the children in generating alternatives
   e. the adult helps the children agree on a solution
   f. the adult reinforces the problem-solving process
   g. the adult aids the children in following through on their agreement

Element: B. By the adult expressing emotions to children

Tasks:

1. the adult identifies the emotions she/he experiences
2. the adult is sensitive to her/his own array of internal cues that signal a particular emotional state
3. the adult uses a wide range of feeling words of different intensities

Element: C. By the adult pinpointing behaviors

Tasks:

1. the adult names the behavior that is affecting her/him
2. the adult describes the behavior, not the child

Element: D. By the adult formulating reasons

Tasks:

1. the adult gives the children specific reasons for why she/he approves or disapproves of their behavior
2. the adult phrases reasons in terms the children understand
3. the adult gives a reason every time she/he attempts to change a child's behavior

Element: E. By the adult formulating rules

Tasks:

1. the adult studies child-development norms
2. the adult gets to know the children in her/his group as individuals
3. the adult thinks about what combinations of knowledge and action the children must carry out to successfully follow a given rule
4. the adult only implements legitimate rules
5. the adult tells the children what the rules are
6. the adult rewards the children's approximations of the rule
7. the adult revises unreasonable rules
8. the adult uses language that is clear and to the point
9. the adult ascertains whether the children have the same understanding of the rule that she/he does

10. when in doubt, the adult assumes that the children have not understood, rather than concluding that they are deliberately breaking the rule
    a. the adult repeats her/his words more slowly and articulates more clearly
    b. the adult rephrases her/his message in simpler, more familiar language and emphasizes key words
    c. the adult restates her/his message using a combination of gestures and words
    d. the adult takes the child into an area where there is less interference from noise and other distractions
    e. the adult emphasizes her/his message using physical prompts such as pictures or objects in combination with gestures
    f. the adult demonstrates what she/he wants by doing it her/himself
11. the adult practices thinking about what she/he wants the children to do as well as what she/he wishes they would refrain from doing
12. the adult catches her/himself saying "No" or "Stop," and rephrases her/his negative instruction as a positive statement
13. the adult tells younger, less experienced children what the alternatives are--the adult lets older or more experienced children generate alternatives for themselves
14. the adult sets consistent limits on children's aggressive behavior

Element: F. By the adult implementing logical consequences when the children break rules

Tasks:
1. the adult anticipates logical consequences that fit the rules she/he makes
2. the adult gives the children opportunities to generate their own ideas for rules and consequences
3. the adult articulates the consequences in the form of a warning

4. the adult gives the children warnings privately
5. the adult points out the natural consequences of children's actions
6. the adult uses the personal message, warning, and follow-through in order
7. the adult allows the children enough time to respond to each step of the sequence
8. the adult finishes the follow-through once she/he begins it
9. the adult communicates with the other adults regarding rule enforcement
10. the adult avoids power struggles
    a. the adult avoids making unnecessary rules
    b. the adult does not embarrass the children in public--she/he keeps all communications between her/himself and the child private
    c. the adult remains calm
    d. the adult avoids contradicting the children's assertions
    e. the adult sticks to the main issue--she/he does not allow her/himself to become involved in an argument over extraneous details
    f. the adult discusses the power struggle privately with the child--this strategy is particularly effective with older children who have learned some attributes of compromise--the adult tells the child directly that a power struggle seems to be developing and that she/he would like to work out the issue in another way
    g. the adult avoids entrapment--when children begin to argue, the adult refuses to become involved--the adult can do this either by quietly repeating the rule and the consequences and then she/he resumes her/his normal activity, or the adult tells the child that she/he would be willing to discuss it later when both the adult and the child are more calm
11. the adult teaches the children self-instructional strategies

12. the adult actively attempts to alter the children's faulty perceptions
13. the adult prepares in advance to use time-out
14. the adult uses time-out only with children who are having a temper tantrum or who exhibit habitual antisocial behavior
15. the adult intervenes immediately in aggressive encounters
16. the adult responds immediately to the children's verbal hostile aggression using the warning and follow through parts of the personal message
17. the adult talks and acts simultaneously to respond to children's unprovoked hostile aggression and to stop children's actions that may be harmful to themselves or others using the warning and follow through part of the personal message
18. the adult attends to the victims of aggression

Function: VI. By Helping Children Understand Stressful Situations in Their Lives and Learn to Use Effective Strategies When Facing Them (such as abuse, death, divorce, sickness, unemployment of a parent, unexpected separation from their parents)

Tasks:

1. the adult uses appropriate vocabulary when discussing death and dying
2. the adult describes death in terms of familiar bodily functions
3. the adult explains why the death has occurred, giving the children accurate information
4. the adult explains death rituals as a means by which people provide comfort to the living
5. the adult answers the children's questions about death matter-of-factly
6. the adult respects the family's prerogative for giving children religious explanations about death
7. the adult explains to the children that divorce is the result of "grown-up problems"
8. the adult acknowledges the pain that divorce inevitably brings to children

9. the adult helps children formulate ways to cope with aggressors beyond the adult's jurisdiction
10. the adult teaches the children specific relaxation techniques
11. the adult helps the children practice imagery

Function: VII. By Helping Children Understand, Accept, and Value Individual Differences and Similarities Including the Sensitive Arenas of Gender Role, Sexuality, Ethnicity, Social Class, and Handicapping Conditions

Tasks:

1. the adult educates her/himself about persons of varying cultural, religious, racial, and developmental backgrounds
2. the adult evaluates her/his own responses to the sensitive areas of sexuality, ethnicity, and handicapping conditions
3. the adult identifies the children who have health related problems or developmental delays
4. the adult responds thoughtfully to children's questions about sexuality, ethnicity, and/or handicapping conditions
5. the adult uses correct vocabulary when referring to body parts, cultural groups, or handicapping conditions
6. the adult reacts calmly to children's sexual play
7. the adult provides natural opportunities for children to learn more about their sexual development
8. the adult helps the children develop appreciation for our diverse heritage as a society
9. the adult utilizes rules and consequences to let the children know that purposeful slurs and unkind references to particular children or groups will not be tolerated
10. the adult monitors all teaching materials and activities for racial, cultural, class, gender role, sexual, religious, and developmental stereotypes
11. the adult respects cultural and experiential differences in children

Function: VIII. By Influencing Children's Social Development Via Structuring the Environment

Element: A. By the adult establishing a daily schedule

1. the adult plans the schedule in detail and writes it down
2. the adult describes the daily schedule to the children
3. the adult tells the children in advance that a transition will occur
4. the adult walks the children through the schedule on their first day or assigns another child to escort them when moving between unfamiliar places
5. the adult sends the children to an activity rather than away from one
6. the adult evaluates the effectiveness of the schedule at regular intervals
7. the adult builds relaxing breaks into the program
8. the adult provides opportunities for vigorous daily exercise

Element: B. By the adult changing the qualities of the room to correct problems in or to further enhance a positive atmosphere of the environment

Tasks:

1. the adult determines whether or not the physical environment supports her/his goals for the children
2. the adult adds or subtracts objects in the physical environment to achieve specific goals related to the children's social development
3. the adult eliminates unnecessary competition
4. the adult eliminates aggressive materials from the setting

Element: C. By the adult effectively managing materials for the children's use

Tasks:

1. the adult stores materials to be used by the children in durable containers near the point of first use and so that they are easy to reach, grasp, and use
2. the adult establishes specific locations for materials so that the children will know where to put them away

3. the adult checks the equipment and the materials to be sure that they are complete, safe, and usable
4. the adult demonstrates the proper use and care of materials
5. the adult gives reasons for the standards that she/he has set
6. the adult supervises the process of putting materials away, giving reminders as necessary, praising the children who are achieving the standard and those who help others to do so

Element: D. By the adult manipulating the environment to minimize conflict

1. the adult provides only enough chairs for the maximum number of children that can participate in an activity
2. the adult encourages children to personalize their space by letting them make room decorations, use the bulletin boards, and have a display area
3. the adult provides for appropriate activities for a private space
4. the adult uses materials that are developmentally appropriate
5. the adult uses furnishings of appropriate size
6. the adult provides materials in an appropriate number for the task and the situation
7. for young children, especially toddlers, the adult provides duplicate or near duplicate play materials
8. the adult arranges the space so that the children can get materials and take care of them without interfering with the other children

Element: E. By the adult helping the children to make decisions and to manage independently

1. the adult offers many different choices to the children each day
2. the adult takes advantage of naturally occurring situations in which to offer choices to the children
3. the adult offers choices to the children using positive statements

4. the adult offers choices to the children for which she/he is willing to accept either alternative the children select
5. the adult allows the children ample time to make their decisions
6. the adult allows the children to change their minds if the follow through on the decisions has not yet begun
7. the adult assists the children in accepting responsibility for the choices they make
8. the adult allows the children, within their developmental abilities, to collaborate with her/him on major management problems such as the storage of materials, proposed rearrangement of space, or planning a special activity
9. the adult allows the children to participate in decision making and conflict resolution
10. the adult allows the children to experience the positive and negative consequences of their decisions unless doing so would endanger their safety, their physical health, or their emotional health

Function: IX. By Supporting Children's Friendships

Element: A. By the adult creating an environment in which children's friendships are respected and encouraged

1. the adult provides opportunities for the children to be with their friends informally--to talk, to play, and to enjoy one another's company
2. the adult plans ways to pair children in order to facilitate interactions
3. the adult pairs a shy child with a younger playmate who is less sophisticated socially
4. the adult takes children's friendships seriously
5. the adult carries out group discussions that focus on children's self-discovered similarities
6. the adult helps the children learn each other's names

7. the adult gives the children on-the-spot information to help them recognize the friendly overtures of others
8. the adult helps the children recognize how their behavior affects their ability to make friends
9. the adult gets the children involved at the beginning of a play episode so that they will not be viewed as interlopers
10. the adult helps the children endure the sorrows of friendships
11. the adult carries out group discussions that highlight friendship-related facts and principles
12. the adult plans how to modify individual differences that cause children problems in their interactions with others
13. the adult designs teaching skits that demonstrate friendship skills to the children
14. the adult encourages the older children to make up skits of their own that dramatize a problem with friends

Element: B. By the adult teaching the children how to role play

Tasks:

1. the adult explains what role playing is
2. the adult sets the scene
3. the adult helps the role players get into character
4. the adult watches the role players attentively
5. the adult discusses what occurred during the role play episode
6. the adult asks the children to develop alternate scenarios
7. the adult summarizes the key points of the children's discussion

Element: C. By the adult carrying out friendship coaching

Tasks:

1. the adult selects a skill to work on
2. the adult initiates coaching
3. the adult describes the skill to the child
4. the adult demonstrates the skill to the child
5. the adult provides a rationale to the child for the skill

6. the adult tells the child to practice the skill
7. the adult evaluates the child's use of the skill
8. the adult repeats the coaching procedure several times

Function: X. By Promoting Responsible Social Behavior

Element: A. By the adult creating a prosocial environment

Tasks:

1. the adult takes advantage of naturally occurring opportunities to label the children's prosocial acts
2. the adult points out instances in which an unintended lack of kindness was shown and describes an alternative, prosocial approach
3. the adult creates opportunities for the children to cooperate
4. the adult creates opportunities for the children to help
5. the adult explains potential or current deviations from rules that are made in order to promote helping responses
6. the adult rewards prosocial behavior
7. the adult administers group rewards
8. the adult models a variety of prosocial behaviors
9. the adult models constructive ways of responding to other people's prosocial behavior
10. the adult is positive when engaging in prosocial behavior
11. the adult points out the prosocial behaviors modeled by her/himself and others
12. the adult builds a positive social climate in which both similarities and differences are valued
13. the adult builds a cooperative, rather than a competitive, spirit within the group

Element: B. By the adult providing direct instruction related to prosocial behavior

Tasks:

1. the adult observes the children for signs of prosocial behavior
2. the adult makes the children aware of when someone needs help

3. the adult teaches the children signals that they might give to elicit help or cooperation from others
4. the adult points out situations in which people could decide to help or cooperate
5. the adult discusses situations in which it would be best to decide not to cooperate
6. the adult assists the children in determining what type of help or cooperation is most suitable for a particular situation
7. the adult works with the children to evaluate the results of their actions
8. the adult encourages the children to accept help from others
9. the adult supports the children when their attempts at kindness are rebuffed
10. the adult uses teaching materials, strategies, and resources that promote divergent as well as convergent thought in the context of social situations
11. the adult learns about what goes on in the children's lives away from the program and takes this into account when planning for the children
12. the adult models nonaggressive behavior
13. the adult points out instances of accidental aggression when they occur
14. the adult uses substitution in response to the children's expressive aggression
15. the adult praises the children when they attempt nonaggressive solutions to difficult situations
16. the adult provides accurate information when children assume that, because society condones aggression in one arena, it is permissible in all arenas
17. the adult points out to the children that individuals can choose nonaggressive solutions to problems
18. the adult uses planned activities to increase children's awareness of alternatives to aggression